

CPI PREDICTORS OF CLINICAL EFFECTIVENESS OF
VOLUNTEERS IN A SUICIDE AND CRISIS
INTERVENTION SERVICE: (I) FACTOR MEASURES
OF STABILITY AND EXTRAVERSION
(II) A CLINICAL EFFECTIVENESS SCALE

By

ROBERT R. BELANGER

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF THE
UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA
1972

Copyright by
ROBERT R. BELANGER
1972

To Dianna

ACKNOWLEDGMENTS

The writer wishes to express his appreciation of the advice and help extended to him by the members of his supervisory committee: Dr. Richard K. McGee, chairman, Dr. Wayne C. Richard, Dr. Madelaine Carey-Ramey, Dr. Harry A. Grater, and a special thanks to Dr. John G. Saw for his statistical consultation.

Also, thanks is given to a good friend, David A. Knickerbocker, for his cooperation and help.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS	iv
LIST OF TABLES	vi
ABSTRACT	viii
INTRODUCTION	1
1. <u>Current Status of Work in the Area</u>	1
2. <u>Rationale of This Approach</u>	27
3. <u>Specific Aims of This Project</u>	28
METHOD	29
<u>Subjects</u>	29
<u>Instruments</u>	29
<u>Procedure</u>	31
<u>Hypotheses</u>	33
RESULTS	35
DISCUSSION	51
APPENDIX	64
BIBLIOGRAPHY	67
BIOGRAPHICAL SKETCH	74

LIST OF TABLES

TABLE	PAGE
1. Factor Loadings of the CPI Scales for Six Studies	13
2. Means and Standard Deviations of All Variables	36
3. Correlation Matrix of All Variables	37
4. Data Pertaining to Hypothesis 1: Comparison of Prediction with Two Factor Measures of Stability (I-V) and Extraversion (II-P) and with the Addition of All the CPI Scales	39
5. Comparison of Multiple Regression and Canonical Correlation	40
6. Items of the Ce-36 Scale and the level of Significance in the Discrimination of the Lowest 10% from the Upper 90% on Clinical Effectiveness (CE)	43
7. Distribution of Ce-36 Scale Scores and Overall Hit Rate for the Discrimination of the Lowest 10% from the Upper 90% on Clinical Effectiveness (CE)	44
8. Comparison of Hit Rates for Ce-36 Scale with Original and Cross-Validation Groups in the Discrimination of the Lowest 10% from the Upper 90% on Clinical Effectiveness (CE)	45
9. Items of the Ce-72 Scale and the Level of Significance in the Discrimination of the Lowest 10% from the Upper 90% on Clinical Effectiveness (CE)	47
10. Distribution of Ce-72 Scale Scores and Overall Hit Rate for the Discrimination of the Lowest 10% from the Upper 90% on Clinical Effectiveness (CE)	48

LIST OF TABLES - Continued

TABLE	PAGE
11. Comparison of CPI Predictors of Clinical Effectiveness (CE): Factor Measures of Stability (I-V) and Extraversion (II-P) Used in Multiple Regression Vs. Ce-72 Scale Discrimination	49
12. Regression and Canonical Beta Weights for Raw Data in the Prediction of (a) Clinical Effectiveness-CE and (b) Genuineness-G, Empathy-E, and Warmth-W	65
13. Multivariate Test of Significance for the Addition of (q) Predictor Variables, to (p) Initial Variables in a Multiple Regression Analysis	66

A Dissertation Presented to the Graduate Council of the
University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

CPI PREDICTORS OF CLINICAL EFFECTIVENESS OF
VOLUNTEERS IN A SUICIDE AND CRISIS
INTERVENTION SERVICE: (I) FACTOR MEASURES
OF STABILITY AND EXTRAVERSION
(II) A CLINICAL EFFECTIVENESS SCALE

By

ROBERT R. BELANGER

March, 1972

Chairman: Dr. Richard K. McGee
Major Department: Psychology

Seventy-two volunteers for telephone duty in a 24-hour suicide intervention service, took the California Psychological Inventory (CPI), and were rated on the Truax and Carkhuff (1967) scales of genuineness, empathy and warmth. (Volunteer data is available through the Volunteer Data Bank of the Suicide and Crisis Intervention Service, Gainesville, Florida.) A composite of these rated measures was described in this study as Clinical Effectiveness (CE) of volunteers.

In Part I, multiple regression was employed using two CPI scales, derived from factor analytic studies, Factor I-V and Factor II-P (respectively, measures of stability and extraversion) in the prediction of CE. It was hypothesized that the other 18 standard CPI scales would not add significantly, beyond the factor measures of

stability and extraversion, to the prediction of CE. This was supported and a further analysis, using canonical correlation in the prediction of genuineness, empathy and warmth, was made for a comparison with the multiple regression. Although the canonical analysis preserves more information and is generally more appropriate, neither the canonical nor the multiple regression method produced a means of predicting the criterion accurately enough for most useful purposes.

In Part II, an attempt was made to develop an Inflexibility scale from a subset of items of the Flexibility (Fx) scale of the CPI, to discriminate the lowest 10% of volunteers from the rest on the criterion of CE. Since only 3 items of the Fx scale were related to the criterion of CE, the remaining items of the CPI were analyzed and a 17-item scale was developed based solely on a random sample of 36 subjects. The scale was designated as the Clinical effectiveness-36 scale (Ce-36) and was cross-validated on a hold-out group. Of 36 subjects in the cross-validation group, 3 of the lowest 4 on CE were correctly classified, while only 2 of the upper 32 were misclassified, using only the Ce-36 scale (significant at the .001 level). A final 19-item scale was developed based on the entire sample of 72 subjects, and was designated as the Clinical effectiveness-72 scale (Ce-72). Eleven items of the Ce-72 scale were identical to the Ce-36 scale. With the Ce-72

scale, 6 of the lowest 7 and 64 of the upper 65 were correctly categorized. It was concluded that the derived scales show some validity for the discrimination of the extreme lowest group on CE. However, there is evidence that the scale may only be applicable to the criterion of CE with volunteers specifically in a suicide and crisis intervention telephone service. Further validation is required to resolve this question.

INTRODUCTION

1. Current Status of Work in the Area

A vast amount of research has been directed at the discovery of personality test correlates of therapeutic or clinical effectiveness in helping relationships. Not only are there an endless array of personality measures possible, there are an equally large number of standards for the evaluation of therapeutic ability. The rationale for the use of the measures employed in this research will be made explicit by a review of the literature in 4 specific areas: (A) A review of studies will be made which indicates a relationship between the ratings of 3 therapy and counseling process variables - genuineness, empathy and warmth - and a wide spectrum of outcome measures. (B) A review will be made of the studies investigating the factorial structure and general use of the California Psychological Inventory (CPI) as a personality measure. (C) Several studies have researched various personality correlates of the ratings of genuineness, empathy and warmth in a variety of settings. (D) Finally, the research in the context of suicide prevention and crisis intervention centers will be reviewed and issues will be considered that are directly relevant

to the circumstances of the present investigation.

(A) Genuineness, Empathy and Warmth Related to Outcome

The major Truax and Carkhuff (1967) work defines the rating scale of genuineness:

This scale is an attempt to define five degrees of therapist genuineness, beginning at a very low level where the therapist presents a facade or defends and denies feelings; and continuing to a high level of self-congruence where the therapist is freely and deeply himself. A high level of self-congruence does not mean that the therapist must overtly express his feelings but only that he does not deny them. Thus, the therapist may be actively reflecting, interpreting, analyzing, or in other ways functioning as a therapist, but this functioning must be self-congruent, so that he is being himself in the moment rather than presenting a professional facade. Thus the therapist's response must be sincere rather than phony; it must express his real feelings or being rather than defensiveness. (1967, pp. 68-69)

Empathy is defined as:

Accurate empathy involves more than just the ability of the therapist to sense the client or patient's "private world" as if it were his own. It also involves more than just his ability to know what the patient means. Accurate empathy involves both the therapist's sensitivity to current feelings and his verbal facility to communicate this understanding in a language attuned to the client's current feelings... It is not necessary - indeed it would seem undesirable - for the therapist to share the client's feelings in any sense that would require him to feel the same emotions. It is instead an appreciation and a sensitive awareness of those feelings. At deeper levels of empathy, it also involves enough understanding of patterns of human feelings and experience to sense feelings that the client only partially reveals. With such experience and knowledge, the therapist can communicate what the client clearly knows as well as meanings in the client's experience of which he is scarcely aware. (1967, p. 46)

Warmth is described:

The dimension of nonpossessive warmth or unconditional positive regard, ranges from a high level where the therapist warmly accepts the patient's experience as part of that person, without imposing conditions; to a low level where the therapist evaluates a patient or his feelings, expresses dislike or disapproval, or expresses warmth in a selective and evaluative way... Thus, a warm positive feeling toward the client may still rate quite low in this scale if it is given conditionally. Nonpossessive warmth for the client means accepting him as a person with human potentialities. It involves a nonpossessive caring for him as a separate person and, thus, a willingness to share equally his joys and aspirations or his depressions and failures. It involves valuing the patient as a person, separate from any evaluation of his behavior or thoughts... The therapist's response to the patient's thoughts or behaviors is a search for their meaning or value within the patient rather than disapproval or approval. (1967, p. 60)

Raters are trained to use these scales in rating tape recordings of interviews, usually three minutes in duration. These ratings are then evaluated for interrater reliability and correlated with various outcome measures.

The results of a large number of studies over the past 13 years have been overwhelmingly in favor of the hypothesis that the ratings are significantly related to outcome. Some of the early landmark research are: Halkides, 1958; Betz, 1963; Rogers, 1962; Truax, 1963; Truax and Carkhuff, 1963. The pattern has continued, and has included some very recent work: Truax, 1968; Truax and Wargo, 1969; Shaw, 1970; Truax, 1970; Garfield and Berzin, 1971; Truax, Wittmer, and Wargo, 1971; Mullen, 1970; Truax and Mitchell, 1971. In compiling data

from this prolific area of psychotherapy and counseling research, a consistent trend evolves which demonstrates that approximately 97% of all outcome measures used, in 42 studies from 1961 to the present, have been significantly related to ratings of genuineness, empathy, and warmth. (Interrater reliabilities range from .23 to .95 with a median of .72 - Truax and Mitchell, 1971). The outcome measures used include MMPI change scores, special scales and indices of the MMPI, Q-sort self ratings, and such behavioral measures as time away from institutions and ratings by friends and acquaintances of clients.

Not only is this research over many years, but it includes differing therapists and theoretical orientations, duration and type of therapy and counseling, type of client and a broad range of measures used to evaluate therapeutic outcome. The client population included "psychotics", "neurotics" and "normals" in group and individual therapy and such diverse categories are represented as: hospital inpatients and outpatients, mental retardates, delinquents; college, high school and elementary students in both counseling and classroom interactions. None of these variables appreciably change the persistent relationship between ratings and outcome. In discussing the effects of diverse sampling, Truax and Carkhuff (1967) state:

The variety of patient populations in different studies were sampled from the South, Midwest, and Northeastern United States and included a variety of socioeconomic, personality, age, sex, and even culture differences. To date, no pattern has emerged, and the findings seem not to change radically from one sample of patients or clients to the next.

Similarly, the findings for the importance of empathy, warmth, and genuineness do not seem to vary radically with the nature of the therapist sample - regardless of theoretic orientation, discipline, age, experience level, geographic location, or even beliefs. (1967, p. 129)

It can be further stated that other types of interactions have been researched which indicate that the therapeutic conditions are meaningful in nontherapy or quasi-therapy situations, e.g., parent-child dyads, studies of mass communication and small group processes, and friends and family relations. Since the outcome of quasi-therapeutic interactions reflect on the generality of the rated conditions, a brief description of these studies is valuable.

Several researchers (cited in Truax and Carkhuff, 1967) have found that when the parent-child relationship can be described as strongly ingenuine, distant and stereotyped or defensive, there is an increased likelihood of emotional disturbance in the child (Baxter, Becker and Hooks, 1963; Bowen, 1960). In a study by Schulman et.al. (1962) it was determined that parents of children (8 to 12 years old) who had conduct problems were significantly more hostile and rejecting toward their children during an observation period. Bishop (1951) observed a high frequency of noncooperative,

reluctant, and inhibited behaviors of children whose mothers lacked warmth or were possessive.

Similar findings have been observed in mass communication and public opinion research. Empathic responsiveness has been highly related with an ability to communicate effectively (Gompertz, 1960). Zimbardo (1960) has shown a relationship between high levels of personal involvement of the communicator in mass communication to a greater degree of opinion change of the recipient. Likewise, opinion change was related to feelings of warmth or liking by Winthrop (1958).

Studies on small group process suggest that groups of high levels of cohesiveness or liking have considerably greater influence over each other (Berkowitz, 1954; Darley et al., 1951). French and Snyder (1959) in analyzing army personnel group interactions, found that persons who were highly liked were more able to influence attitudes and behaviors.

Two recent investigations are particularly relevant to the generality of the effectiveness of the therapeutic conditions, because they employed the Truax and Carkhuff (1967) rating scales and method. Shapiro, Krauss and Truax (1969) found that the disclosure of deeply meaningful personal events between family members and acquaintances was significantly influenced by the level of therapeutic conditions offered. Shapiro and Voog (1969) reported that the college grades of

freshman were significantly related to their roommates level of conditions, i.e., the level of empathy, genuineness and warmth offered by a roommate significantly affected the other's grades. It is of note that in the Shapiro and Voog (1969) study, both positive and hindering effects occurred. In view of this evidence Truax and Mitchell (1971) state that there are those "who habitually help and hurt the people they interact with." (1971, p. 328) Truax and Mitchell (1971) further elaborate:

If the untrained or minimally trained individual has a naturally high level of accurate empathy, nonpossessive warmth, genuineness, and other interpersonal skills, then it seems likely, from the present vantage point, that individuals who spend time with him will be as helped, if not more helped, than if they were receiving formal counseling or psychotherapy from the socially sanctioned professional. (1971, p. 328)

It must be emphasized that these studies are obviously not conclusive, however they do present a pattern whereby genuineness, empathy and warmth may be potentially powerful determiners of a strikingly wide variety of desirable and destructive interpersonal outcomes.

In discussing research both directly or peripherally related to the use of the ratings of the therapeutic conditions, it is pertinent to recognize that the Truax and Carkhuff (1967) research has not gone uncriticized.

Shapiro's (1969) critical review attacks the contention

of the functional independence of the conditions, stating that analyses have not been done which would clearly support the conclusion of either dependence or independence. Chinsky and Rappaport (1970) and Shapiro (1969) have raised questions about the ambiguous theoretical meaning of the scales. They report evidence that the empathy scale does not measure what is intended. The construct validity of the other scales have also been challenged on theoretical grounds (Shapiro, 1969).

Rosenblum (1970) objected to the use of a single interview to rate therapists, stating that this method does not measure the variability of therapists for different clients. In reply to this point, Truax and Mitchell (1971) cite a series of investigations which attempted to resolve the question. In two studies utilizing nontherapy interviews (Truax, 1962; Truax et al., 1966) it was found that therapists were not significantly different on genuineness and empathy but were different on warmth across clients. In two American studies (Truax, 1963; Anderson, 1968) and a cross-cultural investigation using a German translation of the rating scales (Tausch, Eppel, Fittkau and Minsel, 1969) it was shown that, in actual therapy situations, all three of the conditions (including warmth) were not different across clients. From these reports Truax and Mitchell (1971) conclude:

...the evidence is both uniform and strong in indicating that in both therapeutic and informational interviews it is the interviewer, not the patient, who determines what the level of accurate empathy and genuineness will be. Furthermore, considering nonpossessive warmth, the findings from therapy interviews indicate that the different therapists offer different levels of warmth and that patients in general have little affect on the level of warmth offered.... it appears that empathy, warmth and genuineness are relatively permanent personality characteristics and/or attitudes as well as specific interpersonal skills. (1971, p. 325)

Another criticism was made concerning the statistical analyses used in most of the data using the Truax and Carkhuff (1967) rating method. Although standard tests of significance were made with univariate designs, multivariate analyses are required to control adequately for the intercorrelations of outcome measures.

In his critical analysis, Shapiro (1969) has, in contrast, some favorable comments:

In favor of the studies reviewed, it is notable that the important criterion variable of time out of the institution over a follow-up period ranging from 1 to 3½ years, in hospitalized and delinquent populations shows a uniform trend in 10 studies in favor of the hypothesis... It is concluded that there is good evidence that high levels of these conditions is associated with positive therapeutic outcome in a wide variety of patients. (1969, p. 355)

Thus, although issues of methodology and construct validity have been raised, there is little dissent concerning the predictive validity of the conditions.

A final consideration can be given to the possibility

of experimenter bias. Any research which employs ratings is especially susceptible to this kind of distortion. Relative to the studies on the therapeutic conditions, Truax and Carkhuff (1967) address themselves to this question:

Although a number of procedures used in the research were designed to prevent any biasing influence of the investigators (including the use of college students naive to psychotherapy theory and practice as raters, and procedures to prevent the raters from knowing the therapist, the patient or the case outcome), several of the findings could not have been biased by the investigators even if they had tried... (1) In the last six studies completed, the ratings of the levels of therapeutic conditions based on the tape recordings were completed or nearly completed and tabulated before the posttherapy measures were obtained or scored... (2) In almost all the studies, a different investigator from the one collecting the outcome data supervised the raters who rated levels of empathy, warmth, and genuineness. (3) In only two of the studies of institutionalized patients could the institutionalization rate be biased through indirect action by an investigator or by a therapist who believed in the hypothesis - in the other studies, the investigators had no contact with the institution, and a majority of the therapists had voiced their personal disbelief in the hypothesis. (1967, pp. 129-32)

(B) The Use of the California Psychological Inventory (CPI) and Factor Analytic Studies.

The California Psychological Inventory - CPI (Gough, 1957) was developed as a general personality assessment device for use principally with "normal (non-psychiatrically disturbed)" subjects. In his manual (1964) Gough describes the CPI:

The test booklet contains 480 items (12 of these are duplicates, repeated for ease of machine scoring), and yields 18 standard scores... Its scales are addressed principally to personality characteristics important for social living and social interaction. Thus, while it has been found to have special utility with a few problem groups, e.g. persons of delinquent, asocial tendencies, it may be expected to find most general use in schools, colleges, business and industry, and in clinics and counseling agencies whose clientele consists mainly of socially-functioning individuals. (1964, p. 5)

The reliability of the scales range from .46 to .87 with a mean of about .80, and these are considered as acceptable levels for most purposes (Cronbach, 1959). Validity coefficients, though, have been quite low and the discrimination between two extreme groups have been significant but "not impressive". In a critical review in Euros' Fifth Mental Measurements Yearbook (1959), Cronbach states: "One is inclined to say that personality scores with validities in the .20's are worthless, but Gough makes the possibly valid point that one should not expect any one scale to correlate highly with any datum or observed personality" (p. 186). A more favorable review in Buros' Sixth Mental Measurements Yearbook (1965) is made by Kelly: "...there is convincing evidence that each of the scales has some validity when judged against life performance criteria...All in all, however, in this reviewer's opinion this is one of the best if not the best available instruments of its kind" (p. 171).

It might be concluded that although the CPI has poor

predictive and debatable construct validity, it is still considered to be at least as good a personality assessment tool as any of its type.

Factor analytic studies have attempted to simplify and economize the 18 personality scales on the CPI. As expressed by Thorndike (1959), the scales of the CPI, "...provide a very redundant, inefficient and confused picture of individual personalities....Of the 18 scales there are only 4 to fail to correlate at least .50 with some other scale." (p.187). To remedy this situation six major studies have been made in which the CPI was factor analyzed:

(1) Nichols and Schnell (1963) N - 7,670 (Sample composed of mostly college and high school students, male and female).

(2) Mitchell and Jones (1960) N.- 258 (Mostly female student teachers).

(3) Crites, Bechtoldt, Goodstein and Heilbrun (1961) N -124 (College students, male and female).

(4) Pierce - Jones, Mitchell, and King (1962) N - 156 (City school superintendents, mostly male).

(5) Springob and Struening (1964) N - 226 (High school seniors, male and female).

(6) Schludermann and Schludermann (1970) N -177 (Canadian college students, males only).

Table 1
Factor Loadings* of the CPI Scales
For Six Studies

	I						II						III					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
**Do							69,76,76,63,80,71											
Cs							64,59,48,56,72,61						50					
Sy							79,78,50,71,84,82											
Sp							72,62,29,69,75,60						51,40					
Sa							76,77,57,66,84,82											
Wb	74,79,				67,65,72													
Re	68,58,				64,70													
So	57,43,				, 64,70													
Sc	85,92,58,78,92,91																	
To	73,67,				,60,61,81								54,48,45,42					
Gi	76,83,67,75,79,77																	
Cm																		
Ac	77,80,				,57,80,74													
Al	52,47,				, 55,62								56,67,65,54,65,56					
Ie	62,46				,58,46,66								53,50,				, 38	
Pv	77,47,				,45,43,55													
Fx																		
													69,56,55,56,87,61					
Fe																		

*Correlations less than .40 are omitted

**Note: Do-Dominance; Cs-Capacity for Status; Sy-Sociability Sp-Social Presence; Sa-Self Acceptance; Wb-Sense of Well-Being; Re-Responsibility; So-Socialization; Sc-Self Control; To-Tolerance; Gi-Good Impression; Cm-Communality; Ac-Achieve via Conformance; Ai-Achieve via Independence; Ie-Intellectual Efficiency; Py-Psychological Mindedness; Fx-Flexibility; Fe-Femininity.

Analyses were made usually separating males and females; however, Nichols and Schnell (1963) indicate that the factorial structure for the sexes is nearly the same. It can be seen by the comparison of the factor loadings on the first two factors in Table 1, that the studies produced remarkable similarities despite differing geographic and demographic characteristics of the samples, and also differing factor analytic techniques.

In the major analysis and review by Nichols and Schnell (1963) two scales were developed to measure the major factors and they, (in deference to Gough's request) were named Factor I - Value Orientation (Factor I - V), and Factor II - Person Orientation (Factor II - P). Nichols and Schnell (1963) conducted several studies to validate the meaning of the scales. They found that the Factor I - V scale was positively related (at the .01 level) to many other scales and behavior ratings indicating "psychological stability, control and good interpersonal relations.....It correlated negatively with indices of maladjustment and emotionality" (p.332). The Factor II - P scale, also based on several other scales and behavior ratings, correlated with measures of general "activity" level and "outgoingness". They describe this scale as "measuring the familiar Extraversion-Introversion dimension" (p.332). They conclude that "correlations of these scales with a variety of test and behav-

loral variables suggests that they are similar to factors of neuroticism and extraversion found in previous studies" (p.228).

At least one more factor has been extracted from all of the above cited studies. Although no scale has been derived from it, the loadings of CPI scales on this factor are of particular significance, especially in Part II of the present research. This data is also summarized in Table 1. The scales of Flexibility, Achievement via Independence and to a lesser degree Tolerance, all load on Factor III (note that in studies (3) and (4) this factor is Factor IV). The presence of the high loading of the Flexibility scale suggests that this construct may be valuable in the use of the CPI. This issue will be further discussed in the next section.

(C) Personality Correlates of Genuineness Empathy and Warmth

The research on the prediction of the therapeutic conditions of genuineness, empathy and warmth has been comparatively unencouraging. In reviewing the ambiguities of this research literature one is inclined to agree with Carkhuff (1969): "...we can begin where others have ended, with the conclusion of inconclusiveness, and get on with the task at hand, that of developing the selection indexes that will discern those persons who are most capable of (1) making maximum use of training programs and (2) offering

maximum treatment benefits to the distressed persons seeking their help" (p.79). Although such a pessimistic view is understandable it is probably somewhat premature to assume that no personality measures are capable of meaningfully predicting the 3 therapeutic conditions. However, at present there are only meager (but significant) correlations with the conditions reported in the research.

Bergin and Solomon (1963) made an attempt to predict empathy of 13 postinternship students in clinical and counseling psychology, with personality correlates. Both the Pt (psychasthenia) and the D (depression) scales of the MMPI were significantly related to empathy (correlations respectively, $-.54$ and $-.41$). Several scales of the Edwards Personal Preference Scale were also related, e.g., Change, Dominance, Intracception and Order; however, Vesprani (1969) in his cross-validation using 33 untrained female college students as quasi-therapists, failed to support the findings of the earlier study using the EPPS. The MMPI data, though, were supported (Pt, $-.29$; D, $-.34$). Bergin and Jasper (1969) in another replication study added further support of the MMPI scales (Pt, $-.38$; D, $-.31$), but found no agreement of the EPPS scales with the previous two studies. Truax, Silber and Wargo (1966) report a significant relationship for the MMPI Pt and D scales as well as Si (social introversion),

using 16 students in a therapy training program. This study supports the Bergin and Solomon data on the EPPS measure of Order.

In attempting to relate intellectual and achievement variables to overall levels of therapeutic conditions Carkhuff, Piaget and Pierce (1967) found no significant differences between senior psychology majors with high grade-point averages and those with low averages. Bergin and Solomon (1963) found a nonsignificant negative correlation with verbal intelligence and empathy. In one study (Frankel, 1968) the highest 10% on therapeutic conditions were higher than the lowest 10% of 150 college students on the high school grade point average. Also, higher functioning students more frequently listed the helping profession as their vocational choice and were higher on a creative-productive scale. It must be pointed out, though, that the methodology for rating therapeutic conditions in Frankel's research is considerably different from the previous studies and they may not be directly comparable. This issue of variable rating methods similarly makes it difficult to interpret a study (Donnan, Harlan and Thompson, 1969) utilizing Cattell's 16 P.F. Counselors (22) were rated by counselees (880) using the Relationship Inventory (Barret-Lennard, 1959) instead of the usual objective ratings of tape recordings. This Inventory has been related to tape ratings of the 3

conditions and outcome, but the relationship is generally not as predictive (Truax and Carkhuff, 1967). The somewhat ambiguous results show the clearest relationship between the counselors' scores on Factor A scale (warm, sociable vs. reserved, detached) and counselee ratings of warmth. This scale is an introversion-extraversion measure which accounts for the largest amount of variance of any of the scales on the 16 P.F. (Cattell and Eber, 1967). Factor I scale (sensitive, tender-minded vs. realistic, tough-minded) was positively related to genuineness. (In two correlational studies with the MMPI and the 16 P.F. - La Forge, 1960; Cattell and Bolton, 1961 - the Factor A scale was related highest to the S1 scale.)

Two other studies (Tosi, 1970; Cahoon, 1962) using the Barret-Lennard (1959) Relationship Inventory found that for counselor trainees working primarily with clients seeking educational and vocational counseling, the Rokeach Dogmatism scale was negatively related to the counselees ratings of the 3 conditions. Stoffer (1968), in contrast, reported no relationship between the Rokeach Dogmatism scale and objective ratings of taped interviews or outcome measures, for volunteer women in a helping role with elementary children.

In summary, the scales which have been shown to be most clearly related to objective ratings of the therapeutic

conditions are: MMPI - Pt, D, and Si; and the EPPS - Order. The 16 P.F. - Factor A, and the Rokeach Dogmatism scale show a relationship, though the ratings were by counselees using a standard inventory measure, not objective ratings of recordings.

This pattern of results is difficult to interpret because the only test, of these, which has scales based on factor theory is the 16 P.F. and the results of the study using this instrument are ambiguous. Some tentative interpretations can be made by referring to the intercorrelations of these tests in other research.

The MMPI has been correlated with the 16 P.F. (La Forge, 1960; Cattell and Bolton, 1969) and in both studies significant relationships were found between the Pt scale and the O, L, Q₄, Q₃, and C Factors of the 16 P.F. All of these factors are components of the anxiety-stability dimension (Cattell and Eber, 1967). Factor analytic studies of the MMPI also support this finding since the Pt scale has the highest loading on the Welsh First Factor (described as an anxiety measure - Welsh, 1956). The Si scale shows significant correlations with the H, A, and F Factor scales of the 16 P.F., all of which are introversion-extraversion measures. However, the Si scale appears to be more complex than the Pt, since the Factor E (dominance) and Q₂ (self-sufficiency) scales are also significantly related. The

D scale is another complex scale in that both components of anxiety and extraversion, on the 16 P.F., are correlated with it.

The Order scale of the EPPS does not show a highly consistent pattern of correlations, however, Gough (1964) reports in his CPI manual that the highest correlation with EPPS - Order is the CPI Flexibility scale (-.29). Similarly, Rokeach (1960) found correlations from .37 to .55 between the Rokeach Dogmatism Scale and the Gough-Sanford Rigidity Scale (which was an early form of the CPI Flexibility scale scored in reverse). Three other researchers have found significant relationships between measures tangentially related to flexibility and criteria of counseling effectiveness: Rokeach Dogmatism Scale (Stefflre, 1962); psychological openness - a Rorschach measure (Allen, 1967); cognitive flexibility (Sprinthall et al., 1965). These 3 studies used various criteria measures, and for this reason are far from conclusive, though they do suggest that an investigation utilizing a measure of flexibility may be fruitful.

In an attempt to integrate the research evidence one cannot help but be reserved, but a speculative approach is to conceptualize the most consistent and potent predictors of genuineness, empathy and warmth as also closely related to measures of anxiety-stability and introversion-extraversion, and to a lesser extent flexibility.

(D) Research and Issues Relevant to Nonprofessionals in
Suicide and Crisis Intervention Centers

The history of the use of the nonprofessional volunteer has been reviewed elsewhere (Heilig, 1970; Bellak, 1964) and will not be repeated here, other than to mention that the first systematic utilization began with a classic study by Rioch et al. (1960). This project selected and trained housewives to perform a psychotherapeutic function. Based on several criteria the authors indicate a high degree of success, but fail to specify the personality correlates which were employed.

The very rapid growth of suicide and crisis intervention centers has been documented by McGee (1969). Prior to 1965, 11 centers were operating, by 1968 the number had increased to over 90, and the present number is greater than 130 (Resnik, 1970). This growth has been complimented by a commensurate reliance on the nonprofessional. A survey by McGee (1969) in 1969 indicated that 82% of all centers use lay volunteers along with professionals and 25 per cent use only volunteers to answer crisis telephone calls.

Although the proliferation of these centers and the use of volunteers is evident, relatively little experimental research has been produced. An investigation at the WE CARE suicide prevention program in Orlando, Florida (McGee, Pennington and Hegert, 1967) is an important link with the

research cited in the previous section. In this study it was found that the MMPI scales of Si and D correlated highest (-.27 and -.34) with a criterion of length of time with the program. Also, Si correlated (-.50) significantly with the criterion of training course quiz scores. However, neither D or Si was related to judges' (supervisors) ratings of performance. From this evidence it appears that as with the research on the therapeutic conditions the scales D and Si are relatively important predictors of criteria of performance. Furthermore, McGee et al. (1967) found a significant relationship of the Self Acceptance (Sa) and Social Presence (Sp) scales of the CPI with judges' ratings. It can be seen from Table 1 that these scales have two of the highest loadings of the CPI on Factor II - P, a measure of extraversion (Nichols and Schnell, 1963).

Two other studies have been reported which were neither as explicit nor as comprehensive in defining the personality correlates of the volunteers selected. (Heilig, Farberow, Litman and Shneidman, 1968; Resnik, 1968). The Resnik (1968) study does little more than describe the volunteers who were essentially self-selected. According to test and clinical evaluations of the 25 volunteers, only 7 of them were classified as normal with 12 designated as neurotic and 6 psychotic. This group also exhibited a high incidence of suicidal behavior.

The study conducted at the Los Angeles Suicide Prevention Center (Heilig, et al., 1968) placed a considerable emphasis on selection, but their criteria were primarily on an a priori subjective basis. Screening was first accomplished by only considering those who were referred by a local official of the Mental Health Association and members of the staff and professional colleagues. By the further use of personal interviews and a 1,500 word autobiography the final selections were made. (The MMPI was administered although its use was not made explicit.) Of 16 women who applied as volunteers only 6 were excluded. Three were unable to devote the time required, two semi-professionals "wanted immediately to take over the program and direct it" and "one woman was considered unsuitable because of immaturity and low threshold for anxiety" (p. 64). The subjective criteria, then seems to have been based largely on the applicants willingness to accept the limits of the nonprofessional role, and level of maturity and anxiety. In a latter report Pretzel (1970) described (from MMPI data) the selected volunteers:

As a group the volunteers do not try to conform rigidly to rules but instead tend to challenge the conventional and they have both the initiative and the energy to seek new and inventive methods of performing (p. 32).

In summarizing the descriptions, based on personality test correlates which discriminate the "best" from "worst" volunteers, McGee (1967) lists 4 characteristics:

1. they are comfortable with themselves.
2. they love life and live it to the fullest.
3. they experience an active social existence.
4. they have a capacity for influential leadership.

From these characteristics it appears that measures of anxiety-stability and extraversion-introversion may be economical and useful constructs in the description of effective volunteers.

Although evidence from a pilot study by Knickerbocker (1971) is encouraging, no research in suicide and crisis centers has demonstrated that the measured construct of flexibility is related to performance. As Pretzel's (1970) description indicates, this construct from a subjective viewpoint is more plainly a desirable attribute, and other researchers have voiced their agreement. For example, Heilig et al. (1968) in discussing the selection of volunteers state: "Persons judged to be disruptive, hypercritical, complaining or rigid and not able to adapt readily to new situations were excluded early" (p. 3). Likewise, Resnik (1968) in evaluating his sample of volunteers, concludes:

This suggests that personality stability is an important factor, for it allows an individual to handle and tolerate more calls and problems. Good ego strength, as measured by flexibility and problem-solving ability, appeared to be characteristic of the normal workers. For the most part, they handled their cases with minimal solicitation of the consultant's help. The neurotic

and psychotic workers (in that order) were less flexible in their solutions, and more frequently contacted other workers and the consultant for assistance. (p. 438)

A study conducted by Anita Hughes (1969) is especially pertinent to the use of the telephone in crisis centers. It was found that trained raters of empathy (using the Truax Accurate Empathy Scale) did not rate recordings of telephone interviews significantly different from face-to-face interviews, for a given therapist. Although the ratings of this study had low interrater reliability (.53) the results if substantiated have important implications for the generality of rated empathy.

Research in suicide and crisis centers has not in the past employed measures of genuineness, empathy and warmth of volunteers, using the Truax and Carkhuff (1967) scales, though researchers (Heilig, 1968; and McGee, 1969) have described the use of ratings of supervisors. The present research does not assume that the 3 therapeutic conditions are related to outcome in crisis centers. It does intend to define as economically and precisely as possible the relationship between specific personality correlates of volunteers and their levels of the 3 therapeutic conditions. The question of whether these conditions will be related to outcome in the context of crisis centers must remain open until further research. Furthermore, the use of the 3 conditions

as a basis of selection is also an open matter and largely dependent upon the value systems and the ultimate decision process of the administrators of individual centers.

The administrators of a center decide, too, on what selection ratio can be afforded. This is, of course, a function of the available supply of applicants, the number of positions to be filled, the anticipated drop-out rate and the validity of the selection procedures. It has been true in the past for the Gainesville and other centers (Resnik, 1968; McGee, 1967) that it was necessary to accept a large per cent of all applicants. Based on this practical consideration, the present study (Part II) will employ a dichotomization of the lowest 10% from the highest 90% on the criterion of therapeutic conditions. A pilot investigation (Knickerbocker, 1971) has indicated that this lower group is rated at a level which is defined by Truax and Carkhuff (1967) as below minimal standards of functioning. Their pointed conclusion is relevant:

...the various related professions should take an active hand in weeding out or retraining therapists, educators, counselors, etc., who are unable to provide high levels of effective ingredients, and who therefore are likely to provide human encounters that change people for the worse. (p.142)

2. Rationale of This Approach

Several shortcomings are manifest in the research reviewed and the present study will have the advantage of not being encumbered by the following limitations:

(1) No personality test scale based on factor theory has been reported in the prediction of tape recording ratings of genuineness, empathy and warmth in an actual helping relationship. Hence, the results of the past are ambiguous and uneconomical, and predictive personality correlates cannot be described in a maximally simplified manner.

(2) The investigation of personality correlates of the 3 therapeutic conditions has also suffered from the use of a relatively small number of therapists in actual recorded helping relationships. The largest number reported is 36 and most studies have between 6 and 18. The present research will utilize ratings of recordings from 72 volunteers.

(3) No attempt has been made to develop a scale which discriminates those who are extremely low on the 3 conditions. The construct of inflexibility as measured by items of the CPI Flexibility scale is, based on theoretical and empirical evidence, a potential source for such a discriminating measure.

Besides the advantages listed above, this research will present a possible basis for the selection of volunteers in a suicide and crisis intervention service. Whether this basis is utilized is dependent upon subsequent research and rests ultimately with the decision of administrators.

3. Specific Aims of this Approach

The specific aims of this project can be stated as follows:

Part I

(1) To utilize the constructs of stability-anxiety and extraversion-introversion, measured by two CPI scales derived from factor analytic techniques, as economical and efficient personality test correlate predictors of the therapeutic conditions of genuineness, empathy and warmth.

Part II

(2) To investigate the relationship of the construct of inflexibility, as measured by a subset of items on the CPI Flexibility scale, in the discrimination of those who are the extreme lowest on the 3 conditions.

(3) To determine if the scale in (2) can be further improved by an analysis of the remaining items of the CPI.

METHOD

Subjects. All 72 subjects were volunteers for phone duty at the Suicide and Crisis Intervention Service (SCIS) in Gainesville, Florida. Excluded from the sample were all staff members and psychology interns and practicum students on rotation at the SCIS. The sample (30.6% males) includes undergraduate students from the University of Florida, and members of the community of Gainesville, with various occupations. The mean age for the group is 26.6 years.

Instruments. The California Psychological Inventory - CPI - (Gough, 1964) has been described on pages 10-11. The Factor scales of the CPI, I-V and II-P were developed by Nichols and Schnell (1963) and their description of the construction of these scales is of particular interest:

Factor scores for the first two factors were calculated for CPIs from a sample of 300 undergraduate students by summing the standard scores for the representative scales for each factor. These 300 subjects of both sexes had taken the CPI in groups as a part of several previous studies. Factor scales were obtained for both factors using a procedure developed by Webster (1956) for maximizing the correlation of a test with a criterion as follows: The 300 subjects were divided into five groups on the basis of a factor score containing 9%, 20%, 54%, 20%, and 9% of the cases. The number of subjects in each group answering true for each item was tallied

and a D score was obtained for each of the 480 CPI items by weighting the group total, respectively, -2, -1, 0, -1, and -2. This item analysis was done for each factor, and initial scales for the two factors were obtained by selecting all items which were related to a given factor at the .01 level and which were not related to the other factor at the .05 level. This was done in order to keep the correlation between factor scales low. This procedure resulted in a 121-item scale for Factor I and a 55-item scale for Factor II. The procedure described by Webster (1956) for discarding items to maximize the correlation of the scales with the criterion was applied which resulted in the discarding of eight items from Factor Scale I and four items from Factor Scale II. (p. 229)

The Webster (1956) procedure includes those items which have a maximum covariance with a criterion while sequentially eliminating those items which meet a mathematically defined level of redundancy. It is of note, however, that in evaluating the results of utilizing this method, Nichols and Schnell (1963) state: "...this procedure brings little change in cross-validation reliability and factorial validity statistics. In this instance, it did not prove to be worth the considerable labor involved in carrying out the analysis." (p. 230)

An adjusted¹ composite of the ratings of genuineness, accurate empathy, and nonpossessive warmth (Truax and Carkhuff, 1967), for the purposes of this study is described as Clinical Effectiveness (CE).

The 3 rating scales are discussed on pages 2-3. Nonpossessive warmth and genuineness are rated on a 5-point scale and accurate empathy is on a 9-point scale. The CPI scores were obtained at the beginning of the training for the volunteers, and these scores along with the ratings are available through

¹The ratings of empathy were multiplied by 5/9 before being added to the composite. The other ratings were not adjusted.

the Volunteer Data Bank of the Suicide and Crisis Intervention Service, Gainesville Florida.

Procedure. Truax and Carkhuff (1967) describe the rating procedure used in this study. The ratings were available from the Volunteer Data Bank and the method employed was detailed by Knickerbocker¹(1971):

Raters. Twenty-seven undergraduate students in an Introductory Psychology course at the University of Florida volunteered to serve as raters. At the outset of the experiment, they were divided into one of three groups, each of which received instructions in rating one of the variables. The raters were given bonus course credit for their participation. All raters were formally trained in their groups for two hours. This training included learning the procedure, defining the scale to be rated, checking their reliabilities in a group discussion situation, and listening to and rating 10 previously rated standardized three minute tape segments. An effort was made to answer all questions the raters might have. They were then asked to memorize the two methods of rating and to make a commitment to rate as honestly as possible. (It was felt that many of the raters overlearned these rating scales using this procedure). They were then given the following to take home: 1) a practice sheet of written conversations to rate, 2) a copy of Truax's rating stages..

Stimuli. The ninety-two segments of taped telephone conversation, each of three minute duration were used in the experiment. Each segment was taken from the 1st through 4th minute of an initial case call to the SCIS. Tape segments were randomly picked from the library of taped calls stored at the Center for Crisis Intervention Research according to the following criteria: 1) the call had to have been at least six minutes in duration, 2) it had to have been received during the first year of the center's operation (which controlled somewhat for experience factors), and 3) the call had to be judged by the experimenters as involving a crisis of sufficient intensity such that therapeutic conditions would be offered appropriately. (Calls about jobs, information seeking, and about other peoples crises were excluded). A break of one minute

¹The Knickerbocker (1971) study used the same 72 subjects plus 20 additional.

duration was left between tape segments to give the rater ample time to make the required rating. The samples were then randomly placed on four reels of Scotch Magnetic Tape and played to the raters on a Wollensak T-3000 tape recorder.

Ratings. The raters rated the tape segments in two three-hour sessions two weeks apart. Four raters could not attend the second session and were allowed to rate the tapes in an extra session with the experimenter present. Each rater was given a score card with scoring procedures written on it. For each tape segment, the raters first rated according to Truax's stage method... The reliability of all raters combined (for empathy) was .83....(for warmth) was .80...(for genuineness) was .78..as estimated by Ebel's (1951) formula for intraclass correlations. (pp. 8-9)

In Part I, multiple and canonical correlation was used to predict the criteria (see Hypotheses section).

For Part II, a random sample of 36 was chosen from the 72, as a hold-out cross-validation group. CPI items of the Flexibility scale were selected which significantly discriminated between the upper 90% and lower 10% on CE. Tests of significance were carried out using the hypergeometric discrete probability distribution (Lieberman and Owen, 1961). All available subjects were retested with the CPI (for reliability). The remaining items of the CPI which significantly discriminated the two groups were added to determine if discrimination could be improved. Finally, a cross-validation of the resulting scale was made on the randomly selected hold-out group.

Hypotheses. It is hypothesized:

Part I

1. (a) that the constructs of stability-anxiety and extraversion-introversion (as measured respectively by Factor I-V and II-P scales of the CPI) will significantly¹ predict the criterion of volunteer Clinical Effectiveness (CE), using linear multiple regression.
1. (b) that stability-anxiety and extraversion-introversion will account for most of the predictable variation in volunteer CE, i.e., that the other scales of the CPI will not add significantly beyond the two Factor scales, in the prediction of CE.

A comparison of the multiple regression prediction in Hypothesis 1 (a) and (b) will be made with a canonical correlation analysis; i.e., rather than an arbitrary chosen criterion measure of CE, the criterion will be those linear functions, the one of the separate measures of genuineness (G), empathy (E) and warmth (W), the other the CPI scales, which have the largest correlation.

Also, for a comparison with Part II, an attempt will be made to classify the upper 90% and lowest 10% on CE using CPI scales as predictors, utilizing multiple regression techniques.

¹Whenever the term "significant" or "significantly" is used it refers to an alpha level of .05 unless specified otherwise.

Part II

2. (a) that a scale of items (Inflexibility scale) from the Flexibility scale of the CPI, will significantly discriminate the lowest 10% on CE from the rest.
2. (b) that the derived Inflexibility scale will significantly discriminate the cross-validation group.
2. (c) the scale will have an acceptable level of reliability.
3. that any other items of the CPI which are added to the Inflexibility scale will not significantly improve discrimination.
4. that the use of the Inflexibility scale will have a significantly higher hit rate than the use of the CPI scales in Part I, using multiple regression techniques.

RESULTS

Part I

The means and standard deviations of all variables are presented in Table 2 and the matrix of intercorrelations appears in Table 3. This large array of relationships can be simplified by referring to Table 4, which condenses the data to that which is relevant to Hypothesis 1. From this table it can be seen that Hypothesis 1(a) was supported, i.e., the two scales Factor I-V (a measure of stability-anxiety) and Factor II-P (extraversion-introversion) are significantly related to the criterion of Clinical Effectiveness (CE). Hypothesis 1(b) was similarly supported, in that the multivariate test¹ to determine if the 18 CFI scales added to prediction beyond the two Factor scales, was not significant (at the .05 level).

The canonical correlation analysis is compared with the multiple regression technique in Table 5. (The beta weights and the equations for the data in Table 5, is given in Table 12 in the Appendix.)

¹Special thanks is given to Dr. J.G. Saw for communicating this multivariate test. The general form of the test as well as a worked example, based on the data of the present research, appears in the Appendix.

Table 2

Means and Standard Deviations of All Variables

Variable	Mean	Standard Deviation
De 1	30.97	5.23
Cs 2	21.93	3.48
Sy 3	27.28	5.00
Sp 4	39.75	6.24
Sa 5	23.47	4.16
Wb 6	37.08	3.48
Re 7	30.68	4.08
So 8	34.53	4.99
Sc 9	28.61	5.74
To 10	25.32	4.49
Gi 11	18.86	5.35
Cm 12	25.79	1.99
Ac 13	28.54	4.24
AI 14	23.47	3.40
Ie 15	41.85	4.66
Py 16	13.29	2.72
Fx 17	13.04	3.76
Fe 18	20.63	4.06
I-V 19	72.79	12.27
II-P 20	35.29	6.54
CE 21	80.46	13.62
W 22	28.42	5.17
G 23	28.44	5.38
E 24	42.69	9.08

* De-Dominance; Cs-Capacity for Status; Sy-Sociability
 Sp-Social Presence; Sa- Self Acceptance; Wb-Sense of Well-Being
 Re-Responsibility; So-Socialization; Sc-Self Control;
 To-Tolerance; Gi-Good Impression; Cm-Communality; Ac-Achievement
 via Conformance; AI-Achievement via Independence; Ie-Intellectual
 Efficiency; Py-Psychological Mind-ness; Fx-Flexibility; Fe-Femininity;
 I-V -Factor I (stability-anxiety); II-P -Factor II (extraversion-introversion);
 CE-Clinical Effectiveness; W-Warmth; G-Genuineness; E-Empathy.

Table 3

Correlation Matrix of All Variables

Variable:	1	2	3	4	5	6
Do 1		.44	.53	.37	.44	.25
Cs 2			.58	.38	.33	.34
Sy 3				.33	.59	.31
Sp 4					.13	.39
Sa 5						-.10
Wb 6						
Re 7						
So 8						
Sc 9						
Te 10						
G1 11						
Cm 12						

Variable:	13	14	15	16	17	18
Do 1	.42	.36	.54	.28	.04	-.34
Cs 2	.37	.34	.55	.48	.30	-.23
Sy 3	.38	.27	.52	.30	.08	-.30
Sp 4	.28	.43	.48	.36	.39	-.16
Sa 5	.08	.03	.09	.05	.06	-.40
Wb 6	.59	.37	.54	.44	.13	-.09
Re 7	.38	.29	.35	.25	.03	.05
Se 8	.47	.23	.19	.20	-.02	.15
Sc 9	.53	.39	.37	.42	.01	-.05
Te 10	.54	.67	.64	.42	.29	-.23
G1 11	.26	.09	.19	.34	-.17	-.09
Cm 12	.18	.06	-.03	.02	-.07	.17
Ac 13		.38	.59	.40	.06	-.19
A1 14			.52	.47	.53	-.11
Ie 15				.59	.21	-.19
Py 16					.35	-.12
Fx 17						-.23
Fe 18						
I-V 19						
II-P 20						
CE 21						
W 22						
G 23						
E 24						

*See Table 2 for definitions of symbols.

Table 3 -Continued

7	8	9	10	11	12	Variable	
.23	-.01	.01	.34	.13	.07	Do	1
.27	.02	.13	.37	.15	-.07	Cs	2
.21	.10	.15	.44	.23	.11	Sy	3
.10	.12	-.12	.33	.10	.13	Sp	4
-.16	-.20	-.22	.01	-.06	.09	Sa	5
.48	.38	.58	.56	.44	.21	Wb	6
.	.31	.45	.40	.29	.13	Re	7
		.42	.24	.18	.19	So	8
			.60	.56	.04	Sc	9
				.36	-.05	To	10
					-.10	Gi	11
						Cm	12
19	20	21	22	23	24	Variable	
.22	.78	.13	.17	.18	-.03	Do	1
.27	.60	.21	.20	.22	.12	Cs	2
.37	.75	.23	.12	.25	.22	Sy	3
.12	.41	.29	.29	.30	.17	Sp	4
.01	.50	.06	.12	.07	-.03	Sa	5
.60	.34	.30	.34	.19	.26	Wb	6
.42	.20	.23	.12	.27	.20	Re	7
.46	.08	.26	.15	.23	.30	So	8
.68	.06	.16	.05	.17	.20	Sc	9
.52	.33	.25	.13	.25	.27	To	10
.56	.24	.02	.03	.05	.03	Gi	11
.12	.09	.19	.17	.12	.26	Cm	12
.59	.43	.44	.36	.47	.31	Ac	13
.32	.24	.22	.14	.26	.18	Al	14
.45	.50	.25	.17	.31	.18	Ie	15
.40	.23	.15	.08	.16	.14	Py	16
.00	-.04	.12	.05	.13	.13	Fx	17
.11	-.29	-.10	-.14	-.05	-.06	Fe	18
	.32	.22	.10	.28	.20	I-V	19
		.24	.22	.25	.17	II-P	20
			.87	.94	.85	CE	21
				.60	.57	W	22
					.76	G	23
						E	24

*See Table 2 for definitions of symbols.

Table 4

Data Pertaining to Hypothesis 1: Comparison of Prediction
 With Two Factor Measures of Stability (I-V) and Extra-
 version (II-P) and With the Addition of
 All the CPI Scales

Predictors*	Criterion	R	R ²	Sig.
I-V, II-P	CE	.286	.082	.05
I-V, II-P, Do, Cs, Sy, Sp, Sa, Wb, Re, So, Sc, To, G1, Cm, Ac, Al, Ie, Py, Fx, Fe	CE	.538	.289	n.s.
	Increase in R ²		.208	n.s.

*See Table 2 for definitions of symbols.

Table 5

Comparison of Multiple Regression and Canonical Correlation

Predictors	Criterion	R	R _c	R ²	R _c ²	F _p	Red.	T-Red.
(1) I-V, II-P	(a) CE	.286		.082				
	(b) G, E, W		.324 .148		.105 .022	.734 .265	.077 .004	.081
(2) I-V, II-P, Ac	(a) CE	.442		.195				
	(b) G, E, W		.497 .179 .013		.250 .032 .001	.614 .160 .225	.152 .005 .000	.157
(3) I-V, II-P, Ac, Sp	(a) CE	.470		.221				
	(b) G, E, W		.537 .187 .030		.289 .035 .001	.597 .168 .234	.172 .006 .000	.178
(4) I-V, II-P, Ac, Sp, Wb	(a) CE	.471		.222				
	(b) G, E, W		.586 .436 .146		.344 .190 .021	.354 .397 .249	.121 .075 .005	.202
(5) I-V, II-P, Ac, Sp, Wb, Re, Sc, Sc, To, Gl, Cm, Al, Ie, Py, Fx, Fe, Do, Cs, Sy.	(a) CE	.538		.289				
	(b) G, E, W		.738 .599 .522		.546 .360 .274	.097 .262 .651	.048 .097 .177	.318

R - Multiple correlation coefficient; R_c - Canonical correlation coefficient; F_p - Proportion of variance accounted for by a factor; Red. - Redundancy (see text and Cooley and Lohnes, p. 171); T-Red. - Total Redundancy.

From Table 5, the squared multiple correlation coefficient of the linear regression, for the first four comparisons, is only slightly greater than the Total Redundancy¹ measure of the canonical analysis. For example, two predictors account for .082 (R^2) of the variance of the criterion (CE) in the multiple regression, and .081 (Total Redundancy) of the variance of genuineness (G), empathy (E), and warmth (W), in the canonical correlation method. Likewise, for five predictors R^2 is .222 and Total Redundancy is .202. A similar pattern exists for the other comparisons, with the exception of the 20-predictor variable case, where the R^2 (.289) is somewhat less than the Total Redundancy (.318). Thus, generally, it can be concluded that the canonical analysis preserves more information from the original data, while the prediction of the total variation of genuineness, empathy and warmth is nearly as great or greater than the multiple regression prediction using the composite score of Clinical Effectiveness (CE). Refer to the Discussion section for a more detailed consideration of these relationships.

¹Refer to Cooley and Lohnes (1971), pp.171-180.

Part II

Hypotheses 2, 3, and 4 were not supported since only three items from the Flexibility scale were significantly related to rated Clinical Effectiveness, and this is far too few to develop a suitable scale. However, the remaining items of the CPI were analyzed and a 17-item scale was developed, based solely on a random sample of 36 subjects chosen from the original 72. This derived scale is designated as the Clinical effectiveness scale-36 or Ce-36, and the items are listed in Table 6. These items were chosen from the entire pool of 480 CPI items, and only those that significantly (.05 level) discriminated the lowest 10% (in this case 4 subjects) from the upper 90% (32 subjects) on CE were selected. Table 7 shows the distribution of all 36 scale scores and the maximum hit rate for this sample. This scale (Ce-36) was then cross-validated on the randomly selected hold-out group, i.e. the remaining 36 subjects of the original 72. Table 8 facilitates a comparison of the hit rates for the first group and the cross-validation group. In the first group, all 4 of the lowest 10% group were correctly classified and 30 of the upper 32 were placed in the correct category, yielding an overall hit rate of 94.4%. These results were supported with the cross-validation group in that 3 of the lowest 4, and 30 of the upper 32 were correctly classified, giving a hit rate of 91.7%, which

Items of the Ce-35 Scale and the Level of Significance
In the Discrimination of the Lowest 10% from the
Upper 90% on Clinical Effectiveness (CE)

*Item No.	CPI Item	Sig.
1. 340	Our thinking would be a lot better off if we would just forget about words like "probably," "approximately," and "perhaps."	.001
2. 255	Only a fool would try to change our American way of life.	.001
3. 121	I was a slow learner in school.	.01
4. 357	For most questions there is just one right answer, once a person is able to get all the facts.	.01
5. 271	My parents were always very strict and stern with me.	.01
6. 189	In school my marks in deportment were quite regularly bad.	.01
7. 30**	I gossip a little at times.	.05
8. 5	Our thinking would be a lot better off if we would just forget about words like "probably," "approximately," and "perhaps."	.05
9. 214	In school I was sometimes sent to the principal for cutting up.	.05
10. 397	Once I have my mind made up I seldom change it.	.05
11. 471	I sometimes feel that I do not deserve as good a life as I have.	.05
12. 416	I don't think I'm quite as happy as others seem to be.	.05
13. 25	Clever, sarcastic people make me feel very uncomfortable.	.05
14. 217	I think I would like the work of a librarian.	.05
15. 297	At times I have a strong urge to do something harmful or shocking	.05
16. 367	My home life was always very pleasant.	.05
17. 382	Success is a matter of will power.	.05

*The first eleven items also appear on the Ce-72 scale.

**Item no. 30 is scored for a "true" response, all other items are scored for a "false" response.

Table 7

Distribution of Ce-36 Scale Scores and Overall Hit Rate for
the Discrimination of the Lowest 10% From the Upper 90% on

Clinical Effectiveness (CE)

Upper CE Group (90%)

Subject No. Ce-36. Score** Subject No. Ce-36. Score

2	15	31	12
3	17	35	14
5	15	37	16
6	14	39	16
8	14	42	16
9	15	48	13
12	16	49	17
13	15	50	13
14	16	53	15
16	16	57	16
17	14	59	15
19	14	60	15
22	16	63	13
24	13	65	13
26	13	69	17
28	15	72	11

Lowest CE Group (10%)

Overall Hit Rate

= 94.4%

58	7	Predicted*Class	
64	4	Lowest 10%	Upper 90%
68	3		
47	11		
True		Lowest 10%	4
Class		Upper 90%	0
			2
			30

*Ce-36 scale scores less than or equal to 12 are
classified as low.

**Maximum Ce-36 scale score is 17 (17 CPI items)

Table 8

Comparison of Hit Rates for Ce-36 Scale with Original
and Cross-Validation Groups in the Discrimination
of the Lowest 10% from the Upper 90% on Clinical
Effectiveness (CE)

<u>Predictor</u>	<u>Criterion</u>	<u>Predicted* Class</u>		<u>Hit Rate</u>
Ce-36 scale	CE	Lowest 10%	Upper 90%	
(Original 36 Ss)	Lowest 10%	4	0	94.4%
	Upper 90%	2	30	

<u>Predictor</u>	<u>Criterion</u>	<u>Predicted* Class</u>		
Ce-36 scale	CE	Lowest 10%	Upper 90%	
(Cross-Validation)	Lowest 10%	3	1	91.7%**
	Upper 90%	2	30	

*Ce-36 scale scores less than or equal to 12 are classified as low.

**Significant at the .001 level.

is significant at the .001 level (as determined by tables of the hypergeometric distribution).

A final scale (Ce-72) of 19 items was developed in the same way as the previous scale except that it was based on the entire sample of 72 subjects. The items of this scale are found in Table 9, and the distribution of scores and maximum hit rate is presented in Table 10. Eleven of these items also appear in the Ce-36 scale, and from Table 10, the overall hit rate is 97.2%, with 6 of the lowest 7 and 64 of the upper 65 correctly categorized.

The test-retest reliability of the Ce-36 and Ce-72 scales, based on all subjects who were available for retesting (76% of the original 72 subjects) were respectively .46 and .51. These reliabilities must be interpreted with the consideration that most subjects were retested from two to two-and-a-half years after the original testing. Goldberg's (1964) item statistic data suggests that the Ce-36 and Ce-72 scales would have two-week interval test-retest reliability of .65 to .75 (Jones and Goldberg, 1967).

A final comparison between the multiple regression and Ce-72 scale is given for reference in Table 11. The two Factor scales I-V and II-P were employed in multiple prediction and maximum hit rates were generated based on the derived regression equations. Various cut-off

Table 9

Items of the Ce-72 Scale and the Level of Significance
in the Discrimination of the Lowest 10% from the
Upper 90% on Clinical Effectiveness (CE)

*Item No.	CPI Item	Sig.
1-11. 340,255,121,357,271, 189,30**,5,214,397,471	(Listed in Table 6)	
12. 450	A strong person doesn't show his emotions and feelings.	.05
13. 41	For most questions there is just one right answer, once a person is able to get all the facts.	.05
14. 454	My family has objected to the kind of work I do, or plan to do.	.05
15. 76	I often feel as if the world was just passing me by.	.05
16. 1**	I enjoy social gatherings just to be with people.	.05
17. 89	Once a week or oftener I feel suddenly hot all over, without apparent cause.	.05
18. 372	I have reason for feeling jealous of one or more members of my family.	.05
19. 473	I can't really enjoy a rest or vacation unless I have earned it by some hard work.	.05

*The first eleven items also appear on the Ce-36 scale.
**Items no. 30 and 1 are scored for a "true" response,
all other items are scored for a "false" response.

Table 10

Distribution of Ce-72 Scale Scores and Overall Hit Rate for
the Discrimination of the Lowest 10% From the Upper 90% on
Clinical Effectiveness (CE)

Upper CE Group (90%)

<u>S</u> No.	Ce-72**	<u>S</u> No.	Ce-72	<u>S</u> No.	Ce-72	<u>S</u> No.	Ce-72
1	18	18	18	35	16	53	18
2	19	19	17	37	16	54	17
3	19	20	18	38	18	55	17
4	15	22	19	39	16	56	18
5	18	23	19	40	17	57	18
6	18	24	17	41	18	59	19
7	16	25	18	42	18	60	18
8	16	26	16	43	16	61	18
9	18	27	17	44	16	62	17
10	18	28	18	45	19	63	16
12	18	29	17	46	12	65	17
13	19	30	17	48	18	66	18
14	18	31	15	49	19	67	15
15	18	32	18	50	15	69	19
16	17	33	17	51	16	70	18
17	17	34	15	52	19	71	18
						72	14

Lowest CE Group (10%)

21	13
58	11
11	5
36	9
64	6
68	8
47	17

Overall Hit Rate
= 97.2%

Predicted*Class
Lowest 10% Upper 90%

	Lowest 10%	6	1
True Class	Upper 90%	1	64

*Ce-72 scale scores less than or equal to 13 are classified as low.

**Maximum scale score is 19 (19 CPI items).

Table 11
Comparison of CPI Predictors of Clinical Effectiveness (CE):
Factor Measures of Stability (I-V) and Extraversion (II-P)
Used in Multiple Regression Vs. Ce-72 Scale Discrimination

<u>Multiple Regression</u>				<u>Scale Discrimination</u>	
<u>Predictors:</u> I-V, II-P				Ce-72	
		Predicted Class		Predicted Class	
		Lowest 10%	Upper 90%	Lowest 10%	Upper 90%
True Class	Lowest 10%	3	4	3	4
	Upper 90%	6	59	0	65
<u>Hit Rate:</u>		86.1%		94.4%	
True Class	Lowest 10%	4	3	4	3
	Upper 90%	10	55	0	65
<u>Hit Rate:</u>		81.9%		95.8%	
True Class	Lowest 10%	5	2	5	2
	Upper 90%	20	45	0	65
<u>Hit Rate:</u>		65.3%		97.2%	
True Class	Lowest 10%	6	1	6	1
	Upper 90%	48	17	1	64
<u>Hit Rate:</u>		31.9%		97.2%	
True Class	Lowest 10%	7	0	7	0
	Upper 90%	57	8	31	34
<u>Hit Rate:</u>		20.8%		56.9%	

scores were chosen for the Ce-72 scale to facilitate comparisons with the various possible hit rates of the regression method. In every comparison the Ce-72 scale had a higher hit rate than the two Factor scales used jointly in multiple regression. This is especially true when 6 of the lowest 7 are correctly classified, e.g., only 2 misclassifications occur with the Ce-72 scale under this condition (hit rate = 97.2%) but with the multiple regression method 49 misclassifications are observed (hit rate = 31.9%).

DISCUSSION

Part I

The data from the Results suggest that the constructs of stability-anxiety and extraversion-introversion (as measured by the two Factor scales of the CPI) are basic measures in the conceptualization of personality correlates that are predictive of therapeutic effectiveness (measured by Clinical Effectiveness - CE). Such an interpretation must be tempered by the inherent limitations of the present research. There is still a considerable debate about what criterion is optimal in defining therapeutic effectiveness. Even those who would agree that the mounting research evidence is sufficient to demonstrate the usefulness of genuineness, empathy and warmth might emphasize that other variables are also important. Several of these newer rated measures have appeared such as: concreteness or specificity; persuasive potency; client self-exploration (Truax and Carkhuff, 1967); self-disclosure; confrontation; and immediacy (Carkhuff, 1969). All of these variables show some promise but at present only a relatively limited amount of research is available. The problem of defining the best criteria of effectiveness is particularly

difficult in the context of a suicide and crisis intervention center. Various criteria have been proposed which are especially applicable to the volunteers in a telephone crisis center, e.g., Technical Effectiveness or TE (Fowler, 1971). This measure emphasizes, though not exclusively, technical skills which include: maintaining contact, identification of focal problem, evaluation of suicidal potential, assessment of resources, and formulation and mobilization of an action plan (Farberow, et al., 1968). Two other measures which have been considered to be valuable are the amount of time volunteered and total length of service. None of these alternative and complimentary criteria were employed in the present study, and to this extent, general conclusions must be reserved. Yet, it is apparent that supervisors and researchers in suicide and crisis intervention have indicated that traits and attitudes similar to those described by rated genuineness, empathy and warmth are highly valued qualities in a volunteer (Caplan, 1961; Litman et al., 1965; Kaplan and Litman, 1961). So it appears that Clinical Effectiveness (CE) is one of many variables which are considered to be meaningful in evaluating volunteers. It is also clear that the CPI factor measures of stability and extraversion account for most of the predictable variation of CE, i.e., all of the 18 scales do not add significantly to overall prediction beyond the two factor scales. But this finding is limited

by the generally poor, but significant, prediction observed. Only about eight per cent of the variation of CE is predictable from the two factor scales (Table 4) and only 3 of the lowest 7 subjects on CE are correctly classified while 6 of the upper 65 are misclassified, again using only the two scales (Table 11). Thus, an attempt was made, in Part II, to develop a scale to improve prediction, especially for the practical problem of discriminating the lowest ten per cent group on CE from the rest.

Before considering Part II there are two aspects of Part I that require further elaboration: (1) The use of the multivariate test for the multiple regression, (2) the canonical analysis with the use of a redundancy measure.

To determine if several scales significantly increase multiple regression prediction, one must consider all added scales jointly, to control for the probability of rejecting the null hypothesis of no increase, when it is true. (alpha level). In contrast, what is typically done is that each scale is added separately and tested to evaluate a significant increase in prediction and only those scales are retained. The alpha level with such a procedure is only valid if the scales are uncorrelated, but this is rarely the case. Cooley and Lohnes (1971) warn against using such "stepwise regression methods" because of the hazards of capitalization

on chance. Through the efforts of Dr. J.G. Saw, of the University of Florida Statistics Department, a multivariate test was constructed (see Appendix) which evaluates the addition of several predictor variables jointly, thereby controlling the alpha level. An empirical demonstration of the usefulness of the test can be inferred from the data of Table 5. The squared multiple correlation coefficient (R^2) increased from .082 to .195, after the addition of the Achievement via Conformance scale (Ac) to the two scales, Factors I-V and II-P. If this Ac scale were tested separately it would be concluded that it adds significantly to prediction. But when the Ac scale is tested jointly with the other 17 scales, no significant increase is found beyond the two Factor scales. Thus, the multivariate test shows that when 18 scales are added in multiple regression, that increases in prediction of at least the magnitude found with the Ac scale are to be expected by chance alone (with probability .64) using 72 subjects. A spurious interpretation of the data might have been made without this multivariate analysis. Obviously, the test was of value in this case, and would be applicable to most research projects using multiple regression. In this respect, the limitations of the past research with genuineness, empathy, and warmth, related to outcome, is apparent. No attempt to determine

systematically how the three variables jointly relate to outcome has been made. Analyses have been confined to separate univariate tests which indicate that each of the three alone relate significantly to outcome (as does the composite). But, it is not known if each successive variable adds significantly to prediction of outcome, or whether two of the variables add significantly to the third. The multivariate test of this research can help to clarify these and other important questions.

The use of a canonical multivariate analysis is motivated by the same principle that applies to the multivariate test described above, i.e., it is advantageous to analyze jointly variables that are inter-correlated. Since the three criteria, genuineness (G), empathy (E) and warmth (W) were correlated (W and G correlated .60; W and E, .567; G and E, .764), the canonical technique was appropriate. This approach is also more ideal than using the composite of (CE), because of the loss of original information that is necessitated by the summing of variables. Both methods were used for a comparison, and the results are presented in Table 5. The value of R is the usual multiple correlation coefficient which is the product-moment correlation between the linear function of the predictors and the observed value of CE, in a multiple regression equation. The squared multiple correlation

coefficient (R^2) can be described as the proportion of variance of CE that is accounted for by the predictors. The equations for the prediction of CE with various predictors is given for reference in Table 12 of the Appendix. Returning to the data of Table 5, one can compare the multiple regression method with that of the canonical correlation. Comparisons are difficult to make directly, though, because the canonical correlation coefficient (R_C) is the product-moment correlation between a linear function of the predictors and a linear function of G, E and W, rather than a composite (CE). This is further complicated because more than one linear function can be generated, each with its corresponding value of R_C and R_C^2 . The squared canonical correlation coefficient (R_C^2) is the proportion of variance of a statistically defined linear function of G, E and W, that is accounted for by the predictors. This complicated set of relationships is clarified somewhat by a measure proposed by Stewart and Love (1968) and Miller (1969)¹, called redundancy (Red.) and total redundancy (T-Red.). As mentioned above, each R_C value has a corresponding linear function of G, E and W. This linear function can be thought of as a factor and as with any factor the proportion of the variation that it accounts for, of its components, can be computed. For example,

¹See also the discussion by Cooley and Lohnes, 1971, pp. 171-180

from Table 5, line (1) (b), a linear function of G, E and W is generated which maximally correlates with a linear function of the factor scales I-V, and II-P, and this first value of R_c is .324, and R_c^2 is .105. Hence, .105 of the variance of the first linear function of G, E and W is accounted for by the first linear function of I-V and II-P. (The beta weights of these functions are listed in Table 12 of the Appendix). Likewise, only .022 of the variance of the second linear function of G, E and W is accounted for by the second linear function of I-V and II-P. (Every linear function after the first is orthogonal to all corresponding linear functions before it.) Each linear function, or factor accounts, also, for a proportion of variance of the total variation of its components. In this example, the first linear function of G, E and W accounts for .734 of the total variation of G, E and W. The redundancy measure is defined as the product of the proportion of variance of the total variation and the proportion of variance between the predictors and criteria, i.e., the product of .734 and .105 which is .077. The total redundancy (T-Red.) is the sum of all redundancy measures (the sum of .077 and .044 equals .091). This final total redundancy value can be interpreted as the proportion of total variation of G, E and W which is predictable by the scales, I-V and II-P. This can be compared

with the R^2 value of .082, which is the proportion of CE which is predictable by the same two scales. A limitation of this comparison is that no statistical test can be easily formulated because of the complexity of the distributions involved. Nevertheless it can be observed that generally the pattern is consistent: the canonical prediction of total variation of genuineness, empathy and warmth is nearly as great or greater than the multiple regression prediction of the composite score of Clinical Effectiveness. One might contend that if the two methods are nearly equal, in the proportion of variance predicted, why not use the simpler and more familiar method of multiple regression? This conclusion is unwarranted because even though the proportion of variance predicted is nearly the same, a larger pool of variation is being predicted in the canonical analysis and more information is obtained. Therefore, on theoretical and empirical grounds the canonical analysis is more appropriate.

From a more pragmatic point of view, neither the canonical nor the regression method produced a means of prediction or discrimination accurate enough for most useful purposes. This and other considerations led to the analysis of Part II.

Part II

The principal conceptual construct of Part II was flexibility, measured by the Flexibility scale (Fx) of the CPI. Since the standard CPI Fx scale correlated only .115 with CE, it was hypothesized that a scale based on a subset of items from the Fx scale could be developed which was closely related to CE, enough to fulfill the practical need of discriminating the lowest ten per cent of subjects on CE from the rest. As was described in the Results section, this and related hypotheses were not supported because only three items from the Fx scale significantly discriminated the sample. It was concluded, then, that the construct of flexibility, measured from the CPI Fx scale, has only a limited value in the prediction of CE.

In interpreting the derived scales, Ce-36 and Ce-72, one may be inclined, based on the item content (see Tables 6 and 9) to conceptualize many of the items as measuring such constructs as "rigidity," "narrow-mindedness," "dogmatism," "inflexibility," etc. However, this inclination must bend to the more objective research evidence which suggests that the two scales are complex measures of many varied constructs of which flexibility is only one. In fact, nearly all the CPI scale constructs are represented, and no one conceptualization can be

said to be clearly superior. It does appear, though, that the derived scales are meaningfully related to CE, and on this basis were designated as Clinical effectiveness scales, rather than selecting one of the many constructs from the CPI.

The cross-validation of this study was of crucial importance in establishing the Ce-36 scale as a stable and valid discriminator of CE, and from Table 8, the discrimination is far better than would be expected by chance (.001 level). But, one point that is of interest is that the hit rate in the cross-validation sample is 91.7 per cent, and if all subjects were indiscriminantly categorized as belonging to the upper group, then the hit rate would be 90 per cent. At first observation, it may seem that the scale improves the base rate only slightly and, therefore, is of little value. The principles of decision theory (Bross, 1953; Hays, 1963) clarify this situation, but a detailed analysis is not feasible because it would require the assignment of a value or loss function to all possible events and this is difficult and at times somewhat arbitrary. What can be said is that the value one would assign to correctly classifying three of the lowest four subjects on Clinical Effectiveness is relatively high compared to the value of assigning all subjects categorically in the upper group, which means that it is highly desir--

able to identify those who are low on CE, even though some (in this case, two) of the upper group are misclassified.

Although this principle is obvious, it might have been overlooked if only the hit rate was used to evaluate the findings. This emphasizes the advantage of considering the principles of decision theory, at least on an intuitive level, when interpreting data, especially when the discrimination of an extreme group is involved.

The Ce-36 scale was shown to significantly discriminate the cross-validation group of this study, but it is considered imperative that further validation studies be undertaken. The Orlando "We Care" research project (McGee, et. al., 1967) provided the data for a further attempt to show a general validation of the Ce-36 scale. The 22 volunteers in that research project had taken the CPI and were evaluated by supervisors, on several dimensions of performance and an overall performance score. Only about 2 volunteers would be expected to fall in the lowest ten per cent group, but no correspondence was found between those who were low on the overall score or component dimensions, and those who scored lowest on the Ce-36 or Ce-72 scale. It is not possible, at this time, to determine if this lack

of correspondence is a function of the Clinical-effectiveness scales or because the ratings of the "We Care" volunteers were of a different nature than that of CE. It should be pointed out that the ratings of supervisors were not based on actual recorded phone performance, and this may have been the essential difference. The important point to make is that the Ce-36 and Ce-72 scales were developed and validated on the criterion of Clinical Effectiveness (CE) and if other criteria are used the scales may not be meaningful predictors of performance. Nor can it be assured that the scales will be useful in settings where the volunteer population is vastly different from the sample used in this research. Such questions can only be answered by further investigations and the continuing process of cross-validation.

On at least one dimension the sample studied is clearly different from most groups researched in the past, and that is the mean rated level of CE, or as described in other studies, overall facilitative conditions. The mean CE for the Gainesville sample was 2.96, and although this is slightly below the median of a five-point scale, it is considerably higher than that found in other studies of both professional and non-professional helpers (Truax and Carkhuff, 1967).

Whether this indicates that volunteers to a crisis center are generally higher in their level of functioning or that only the specific sample was unusual cannot as yet be resolved, but if this finding is replicated with volunteers in other helping agencies, it would suggest that as a group, they have a rich potential of therapeutic skill, and are well worth selecting and training as members of a comprehensive mental health team. Furthermore, if the Ce scales developed in this study prove to apply similarly to other settings, then a feasible and efficient means may be available of "weeding out" from helping roles those who, based on much empirical evidence, are most likely to be destructive in their interpersonal encounters, and selecting those who are most likely to benefit from a program of training.

APPENDIX

Prediction of (a) Clinical Effectiveness-CE and (b)

Genuineness-G, Empathy-E, and Warmth-W*

$$\begin{array}{l}
1. \text{ (a) } .18(\text{I-V}) + .39(\text{II-P}) + 53.48 = (\text{CE}) \\
\text{ (b) } -.67(\text{I-V}) + .56(\text{II-P}) = -.01(\text{W}) -1.10(\text{G}) + .15(\text{E}) \\
.81(\text{I-V}) - .90(\text{II-P}) = -1.25(\text{W}) + .35(\text{G}) + .56(\text{E}) \\
2. \text{ (a) } -.06(\text{I-V}) + .14(\text{II-P}) + 1.41(\text{Ac}) + 39.74 = (\text{CE}) \\
\text{ (b) } -.15(\text{I-V}) + .15(\text{II-P}) + 1.01(\text{Ac}) = .36(\text{W}) + 1.00(\text{G}) - .35(\text{E}) \\
-1.20(\text{I-V}) + .31(\text{II-P}) + .42(\text{Ac}) = 1.20(\text{W}) - .60(\text{G}) - .56(\text{E}) \\
-.26(\text{I-V}) + 1.14(\text{II-P}) - 1.44(\text{Ac}) = -.25(\text{W}) -1.06(\text{G}) + .70(\text{E}) \\
3. \text{ (a) } -.04(\text{I-V}) + .01(\text{II-P}) + 1.30(\text{Ac}) + .39(\text{Sp}) + 30.50 = (\text{CE}) \\
\text{ (b) } -.11(\text{I-V}) - .01(\text{II-P}) + .86(\text{Ac}) \\
+ .42(\text{Sp}) = +.41(\text{W}) + 1.00(\text{G}) - .41(\text{E}) \\
-1.09(\text{I-V}) + .16(\text{II-P}) + .21(\text{Ac}) \\
+ .31(\text{Sp}) = 1.17(\text{W}) - .50(\text{G}) - .61(\text{E}) \\
.38(\text{I-V}) - .91(\text{II-P}) - .32(\text{Ac}) \\
+ .89(\text{Sp}) = .32(\text{W}) + 1.15(\text{G}) - 1.40(\text{E}) \\
4. \text{ (a) } -.05(\text{I-V}) + .01(\text{II-P}) + 1.27(\text{Ac}) + .38(\text{Sp}) \\
+ .11(\text{Wb}) + 28.77 = (\text{CE}) \\
\text{ (b) } -.37(\text{I-V}) + .07(\text{II-P}) - .90(\text{Ac}) \\
- .56(\text{Sp}) + .83(\text{Wb}) \\
= +.12(\text{W}) - 1.48(\text{G}) + .69(\text{E}) \\
-.82(\text{I-V}) + .11(\text{II-P}) + .26(\text{Ac}) \\
+ .01(\text{Sp}) + 1.10(\text{Wb}) \\
= 1.10(\text{W}) - .62(\text{G}) + .35(\text{E}) \\
+.82(\text{I-V}) - .01(\text{II-P}) - .20(\text{Ac}) \\
- .49(\text{Sp}) + .37(\text{Wb}) = -.64(\text{W}) - .37(\text{G}) + 1.38(\text{E}) \\
5. \text{ (a) } -.03(\text{I-V}) + .55(\text{II-P}) + 1.30(\text{Ac}) + .41(\text{Sp}) - .04(\text{Wb}) - .02(\text{Sy}) \\
- .18(\text{Ie}) + .06(\text{So}) + .15(\text{To}) - .14(\text{Cs}) + .47(\text{Cm}) + .17(\text{A1}) \\
+ .47(\text{Re}) - .03(\text{Py}) - .08(\text{Sc}) - .89(\text{Do}) - .02(\text{Fx}) - .23(\text{Fe}) \\
+ .06(\text{Sa}) - .43(\text{G1}) - 8.23 = (\text{CE}) \\
\text{ (b) } +.32(\text{I-V}) - .83(\text{II-P}) + .71(\text{Ac}) + .57(\text{Sp}) - .57(\text{Wb}) - .04(\text{Sy}) \\
+ .26(\text{Ie}) - .34(\text{So}) - .77(\text{To}) + .22(\text{Cs}) - .26(\text{Cm}) + .23(\text{A1}) \\
+ .18(\text{Re}) - .51(\text{Py}) + .22(\text{Sc}) + .67(\text{Do}) - .00(\text{Fx}) + .32(\text{Fe}) \\
+ .26(\text{Sa}) + .34(\text{G1}) = .07(\text{W}) - 1.45(\text{G}) - 1.43(\text{E}) \\
- .37(\text{I-V}) - .70(\text{II-P}) - .26(\text{Ac}) - .16(\text{Sp}) + 1.00(\text{Wb}) - .42(\text{Sy}) \\
- .10(\text{Ie}) + .04(\text{So}) - .30(\text{To}) + .47(\text{Cs}) - .04(\text{Cm}) + .15(\text{A1}) \\
- .40(\text{Re}) - .24(\text{Py}) - .20(\text{Sc}) + .71(\text{Do}) - .28(\text{Fx}) - .10(\text{Fe}) \\
+ .31(\text{Sa}) + .35(\text{G1}) = +.97(\text{W}) - .74(\text{G}) - .60(\text{E}) \\
+ .29(\text{I-V}) - .35(\text{II-P}) - .60(\text{Ac}) - .22(\text{Sp}) - .58(\text{Wb}) + .22(\text{Sy}) \\
+ .23(\text{Ie}) - .12(\text{So}) - .09(\text{To}) - .13(\text{Cs}) - .17(\text{Cm}) - .13(\text{A1}) \\
- .06(\text{Re}) + .04(\text{Py}) + .21(\text{Sc}) + .48(\text{Do}) + .15(\text{Fx}) + .25(\text{Fe}) \\
- .15(\text{Sa}) + .16(\text{G1}) = -.83(\text{W}) + .06(\text{G}) - .32(\text{E})
\end{array}$$

*See Tables 1 and 2 for definitions of symbols.

Table 13

Multivariate Test** of Significance for the Addition
of (q) Predictor Variables, to (p) Initial
Variables in a Multiple Regression Analysis

* R^2 = multiple correlation coefficient squared (population value)
p = number of initial predictor variables
q = number of added predictor variables
N = number of subjects

$$H_0: R_{p+q}^2 = R_p^2 \quad (\text{sample estimate of } R^2 = R^2)$$

$$H_a: R_{p+q}^2 \neq R_p^2$$

$$\frac{1/q}{1/N-1-p-q} \times \frac{R_{p+q}^2 - R_p^2}{1 - R_{p+q}^2} \sim F_{q, N-1-p-q}$$

Accept H_0 if F is less than $F_{1-\alpha, q, N-p-q-1}$

Reject H_0 if F is greater than "

Example From the Present Research:

$$H_0: R_{20}^2 = R_2^2 \quad R_{20}^2 = .2392 (\text{sample estimate})$$

$$H_a: R_{20}^2 \neq R_2^2 \quad R_2^2 = .0815 \quad " \quad "$$

$$N = 72 \quad P = 2 \quad q = 18 \quad \alpha = .05$$

$$\frac{1/18}{1/51} \times \frac{.2392 - .0815}{1.00 - .2392} = .827 = F$$

$$F_{1-\alpha, q, N-1-p-q} = 1.81$$

Since .827 is less than 1.81

∴ Accept H_0

**Communicated by Dr. J.G. Saw, of the University of Florida
Department of Statistics.

BIBLIOGRAPHY

- Allen, T.W. The effectiveness of counselor trainees as a function of psychological openness. Journal of Counseling Psychology, 1967, 35-40.
- Anderson, D.L. A study of selected dimensions of the counselor client relationship. Unpublished doctoral dissertation, University of North Dakota, 1968.
- Barrett-Lennard, G.T. Dimensions of the client's experience of his therapist associated with personality change. Genetic Psychological Monographs, 1959, No. 43.
- Baxter, J.C., Becker, J., and Hooks, W. Defensive style in the families of schizophrenics and controls. Journal of Abnormal and Social Psychology, 1963, 512-518.
- Bellak, L. (Ed.) Handbook of Community Psychiatry and Community Mental Health. New York: Grune and Stratton, 1964.
- Bergin, A.E., and Jasper, L.G. Correlates of empathy in psychotherapy: A replication. Journal of Abnormal Psychology, 1969, 477-481.
- Bergin, A.E., and Solomon, S. Personality and performance correlates of empathic understanding in psychotherapy. In T. Tomlinson and J. Hart, (Eds.), New directions in client-centered therapy. Boston: Houghton-Mifflin, 1963.
- Berkowitz, L. Group standards, cohesiveness, and productivity. Human Relations, 1954, 509-519.
- Betz, Barbara J. Differential success rates of psychotherapists with "process" and "non-process" schizophrenic patients. American Journal of Psychiatry, 1963, 1090-1091.
- Bishop, B.M. Mother-child interaction and the social behavior of children. Psychological Monographs, 1951, 328.
- Bowen, M. A family concept of schizophrenia. In D. Jackson (Ed.), The etiology of schizophrenics. New York: Basic Books, 1960.

- Bross, I. Design for Decision. New York: Macmillan, 1953.
- Cahoon, R.A. Some counselor attitudes and characteristics related to the counseling relationship. Unpublished doctoral dissertation, Ohio State University, 1962.
- Caplan, H. Prevention of mental disorders in children. New York: Basic Books, 1961.
- Carkhuff, R.R. Helping and human relations. New York: Holt, Rinehart and Winston, Inc., 1969.
- Carkhuff, R.R., Piaget, G., and Pierce, R. The development of skills in interpersonal functioning. Counselor Education and Supervision, 1967, 102-106.
- Cattell, R.B., and Bolton, L.S. What pathological dimensions lie beyond the normal dimensions of the 16 PF, Journal of Consulting and Clinical Psychology, 1961, 18-29.
- Cattell, R.B., and Eber, G.W. Handbook for the sixteen personality factor questionnaire. Champaign, Ill.: Institute for Personality and Ability Testing, 1967.
- Chinsky, J.M., and Rappaport, J. Brief critique of the meaning and reliability of accurate empathy ratings. Psychological Bulletin, 1970, 379-382.
- Cooley, W.W., and Lohnes, P.R. Multivariate data analysis. New York: John Wiley and Sons, Inc., 1971.
- Crites, J.O., Bechtoldt, H.P., Goodstein, L.D. and Heilbrun, A.B. A factor analysis of the California Psychological Inventory. Journal of Applied Psychology, 1961, 408-414.
- Cronbach, L.J. In O. Buros (Ed.), Fifth mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1959, 186.
- Dahlstrom, W.G., and Welsh, G.S. An MMPI handbook. Minneapolis: The University of Minnesota Press, 1960.
- Darley, F.G., et al. Studies of group behavior. Journal of Abnormal and Social Psychology, 1951, 564-576.
- Donnan, G.G., Harlan, G.E. and Thompson, S.A. Counselor personality and level of functioning as perceived by counselees. Journal of Counseling Psychology, 1969, 482-485.
- Ebel, R.D. Estimation of the reliability of raters. Psychometrika, 1951, 407-424.

- Farberow, N.L., Heilig, S.M. and Litman, R.E. Techniques in crisis intervention: a training manual. Los Angeles: Suicide Prevention Center, Inc., 1968.
- Fowler, D.E. Assessing the performance of telephone crisis workers: technical effectiveness. Unpublished research, University of Florida, 1971.
- Frankel, S. The biographic and demographic characteristics of high and low discriminators and communicators. Unpublished research, State University of New York at Buffalo, 1968.
- French, J.R.P., and Snyder, R. Leadership in interpersonal power. In D. Cartwright (Ed.), Studies in social power. Ann Arbor: University of Michigan Press, 1959.
- Garfield, S.L., and Bergin, A.E. Therapeutic conditions and outcome. Journal of Abnormal Psychology, 1971, 108-114.
- Goldberg, L.R. Test-Retest item statistics for the California Psychological Inventory. ORI Research Monograph. Vol. 4, No. 1, 1964.
- Gompertz, K. The relation of empathy to effective communication. Journal Quarterly, 1960, 533-546.
- Gough, H.G. Manual for the California Psychological Inventory. Palo Alto, Calif.: Consulting Psychologists Press, 1957, 1964.
- Halkides, G. An investigation of therapeutic success as a function of four variables. Unpublished doctoral dissertation, University of Chicago, 1958.
- Hays, W.L. Statistics for Psychologist. New York: Holt, Rinehart and Winston, 1963.
- Heilig, S.M., Farberow, N.L., Litman, R.E. and Shneidman, E.S. The role of nonprofessional volunteers in a suicide prevention center. Community Mental Health Journal, 1968, 1-9.
- Hughes, A.E. Study I: the accurate empathy ratings of therapists in telephone and face-to-face interviews. Study II: the effect of group sensitivity-training on the accurate empathy ratings of therapists. Unpublished doctoral dissertation. Ohio State University, 1969.
- Jones, R.R., and Goldberg, L.R. Interrelationships among personality scale parameters: item response stability and scale reliability. Educational and Psychological

Measurement, 1967, 323-333.

- Kaplan, H.N., and Litman, R.E. Telephone appraisal of 100 suicidal emergencies. American Journal of Psychology, 1961, 591-599.
- Kelly, E.L. In O. Buros (Ed.), Sixth mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1965.
- Knickerbocker, D. Clinical effectiveness of nonprofessional and professional telephone workers in a crisis intervention center. Unpublished research, University of Florida, 1971.
- LaForge, R. The MMPI and Cattell 16 PF. Journal of Consulting and Clinical Psychology, 1960, 402-409.
- Lieberman, G.J., and Owen, D.B. Tables of the hypergeometric probability distribution. Stanford, California: Stanford University Press, 1961.
- Litman, R.E., Farberow, N.L., Shneidman, E.W., Heilig, S.M., and Kramer, J.A. Suicide-prevention telephone service. Journal of the American Medical Association, 1965, 111.
- McGee, R.K. U.S. suicide prevention center survey. Unpublished research, University of Florida, 1969.
- McGee, R.K., Pennington, J.C. and Hegert, T. Criteria for selecting and evaluating crisis workers. Unpublished research, University of Florida, 1967.
- Miller, J.K. The development and application of bi-multivariate correlation: a measure of statistical association between multivariate measurement sets. Unpublished dissertation, State University of New York at Buffalo, 1969.
- Mitchell, J.V., and Pierce-Jones, J. A factor analysis of Gough's California Psychological Inventory. Journal of Consulting Psychology, 1960, 453-456.
- Morrison, D.F. Multivariate statistical methods. New York: McGraw-Hill, 1967.
- Mullen, J.A. An investigation of liking in therapy and its relation to the variables of outcome, empathy and therapist experience. Dissertation Abstracts, 1970 (3-B), 1546-1547.
- Nichols, R.O., and Schnell, R.R. Factor scales for the California Psychological Inventory. Journal of Consulting Psychology, 1963, 228-235.

- Pierce-Jones, J., Mitchell, J.V. and King, R. A comparison of the factorial composition of the California Psychological Inventory. Journal of Experimental Education, 1962, 66-70.
- Pretzel, P.W. The volunteer clinical worker at the suicide prevention center. Bulletin of Suicidology, 1970, 29-34.
- Resnik, H.L. (Ed.) Suicidal behaviors: Diagnosis and Management. Boston: Little, Brown and Co., 1968.
- Resnik, H.L. Editorial. Bulletin of Suicidology, 1970, 3-4.
- Rioch, Margaret J., Elkes, C., Flint, A.A., Udansky, G.S., Newman, R.G., and Silber, E. National Institute of Mental Health pilot study in training mental health counselors. American Journal of Orthopsychiatry, 1963, 678-689.
- Rogers, C.R. The interpersonal relationship: The core of guidance. Harvard Educational Review, 1962, 416-429.
- Rokeach, Milton. The open and closed mind. New York: Basic Books, 1960.
- Rosenblum, G. Book Reviews. Community Mental Health Journal, 1970, 465-467.
- Schludermann, Shirin, and Schludermann, E. CPI factor analysis. Journal of Psychology, 1970, 43-50.
- Schulman, R.F., Shoemaker, D.J., and Moelis, I. Laboratory measurements of parental behavior. Journal of Consulting Psychology, 1962, 109-114.
- Shapiro, D.A. Empathy, warmth and genuineness in psychotherapy. British Journal of Social and Clinical Psychology, 1969, 350-361.
- Shapiro, H.G., and Voog, T. Effect of the inherently helpful person on student academic achievement. Journal of Counseling Psychology, 1969, 295-299.
- Shapiro, J.G., Krauss, H.H., and Truax, C.B. Therapeutic conditions and disclosure beyond the therapeutic encounter. Journal of Counseling Psychology, 1969, 290-294.
- Shaw, Sarah F. Empathy and its relationship to selected criteria of counselor effectiveness. Dissertation Abstracts, 1970 (1-A), 163.

- Shneidman, E.S., Farberow, N.L., and Litman, R.E. (Eds.), The psychology of suicide. New York: Basic Books, 1970.
- Springob, H.K., and Struening, E.L. Analysis of the California Psychological Inventory. Journal of Counseling Psychology, 1964, 173-179.
- Sprinthall, N.A., Whitely, J.M., and Mosher, R.L. Cognitive flexibility as a dimension of counselor behavior: Definition, rationale, research, and focus. Unpublished research, Harvard University, 1965.
- Stefflre, B., King, P., and Leafgren, F. Characteristics of counselors judged effective by their peers. Journal of Counseling Psychology, 1962, 335-340.
- Stewart, D.K., and Love, W.A. A general canonical correlation index. Psychological Bulletin, 1968, 160-163.
- Stoffer, D.L. An investigation of positive behavioral change as a function of genuineness, non-possessive warmth and empathic understanding. Unpublished doctoral dissertation, Ohio State University, 1968.
- Tausch, R., Eppel, H., Fittkau, B., and Minsel, R. Variablen und zusammenhänge in der gesprächspsychotherapie. Zeitschrift Für Psychologie, 1969, 93-102. (Cited in Truax and Mitchell, 1971, 330.)
- Thorndike, R.L. In O. Buros (Ed.), Fifth mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1959, 187.
- Tosi, Donald J. Dogmatism within the counselor-client dyad. Journal of Counseling Psychology, 1970, 284-288.
- Truax, C.B. Variations in levels of accurate empathy offered in the psychotherapy relationship and case outcome. Brief Research Reports, Wisconsin Psychiatric Institute, University of Wisconsin, 1962.
- Truax, C.B. Effective ingredients in psychotherapy: An approach to unraveling the patient-therapist interaction. Journal of Counseling Psychology, 1963, 256-263.
- Truax, C.B. Therapist interpersonal reinforcement of client self-exploration and therapeutic outcome in group psychotherapy. Journal of Counseling Psychology, 1968, 225-231.

- Truax, C.B. Length of therapist responses, accurate empathy and patient improvement. Journal of Clinical Psychology, 1970, 539.
- Truax, C.B., and Carkhuff, R.R. For better or for worse: The process of psychotherapeutic personality change. Chapter in Recent advances in the study of behavior change. Montreal, Canada: McGill University Press, 1963.
- Truax, C.B., and Carkhuff, R.R. Toward effective counseling and psychotherapy: Training and practice. Chicago: Aldine, 1967.
- Truax, C.B., and Mitchell, K.M. Research on certain therapist interpersonal skills in relation to process and outcome. In A.E. Bergin, and S.L. Garfield (Eds.), Handbook of psychotherapy and behavior change. New York: John Wiley and Sons, 1971.
- Truax, C.B., and Silber, L.D. and Wargo, D.G. Training and change in psychotherapeutic skills. Unpublished manuscript, University of Arkansas, 1966.
- Truax, C.B., and Wargo, D.G. Antecedents to outcome in group psychotherapy with outpatients: Effects of therapeutic conditions, alternate sessions, vicarious therapy pre-training and patient self-exploration. Journal of Consulting and Clinical Psychology, 1969, 122-129.
- Truax, C.B., Wittmer, J., and Wargo, D.G. Effects of the therapeutic conditions of accurate empathy, non-possessive warmth and genuineness on hospitalized mental patients during group therapy. Journal of Clinical Psychology, 1971, 137-142.
- Vesprani, G.J. Personality correlates of accurate empathy in a college companion program. Doctoral dissertation, University of Cincinnati, 1967.
- Webster, H. Maximizing test validity by item selection. Psychometrika, 1956, 153-164.
- Welsh, G.S. Factor dimensions A and R. In G.S. Welsh and W.G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: University of Minnesota Press, 1956.
- Winthrop, H. Relation between appeal value and highbrow status on some radio and television programs. Psychological Reports, 1958, 53-54.
- Zimbardo, P.G. Involvement and communication discrepancy as determinants of opinion conformity. Journal of Abnormal and Social Psychology, 1960, 86-94.

BIOGRAPHICAL SKETCH

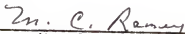
Robert R. Belanger was born June 6, 1945, at Barre, Vermont. In June, 1963, he was graduated from Bishop Amat High School in La Puente, California. After attending two years at Citrus Junior College in Azusa, California, he transferred to California State College at Los Angeles, where he received the degree of Bachelor of Arts with a major in psychology, September, 1967. He began his graduate study in psychology at California State College October, 1967, and transferred to the Graduate School of the University of Florida January, 1968. He received the Master of Arts degree in psychology, December 1968. He plans to pursue a career in clinical psychology.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



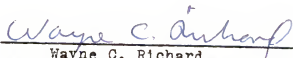
Richard K. McGee, Chairman
Associate Professor of Clinical
Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



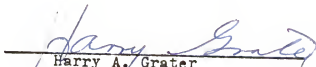
Madelaine Carey-Ramey
Assistant Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Wayne C. Richard
Assistant Professor of Clinical
Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Harry A. Grater
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

J. G. Saw

John G. Saw
Professor of Statistics

This dissertation was submitted to the Department of Psychology in the College of Arts and Sciences and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

March, 1972

Dean, Graduate School